Immersive Virtual Reality: Analyzing the impact of user characteristics

Man Vien, Santa Ana College
Dr. Doina Bein, Department of Computer Science, California State University, Fullerton

Background

- Virtual Reality (VR) technology has revolutionized digital interactions, offering immersive experiences beyond traditional screens by simulating artificial environments that users can interact with and be present in. As VR technology continues to evolve, understanding and enhancing user experiences within virtual environments become crucial. Our research aims to answer the following question: How do demographic characteristics, VR headset type, duration of experience, and motion sickness rating influence the immersion level of users in virtual reality environments? Using machine learning to analyze the available dataset, we aim to uncover patterns and correlations that enhance our understanding of the relationships of these variables. This will allow us to make predictions on the immersion level based on these variables and gain valuable insights for optimizing and personalizing VR experiences.
- Machine learning, a subset of artificial intelligence, enables computer systems to learn from data and make predictions based on patterns and relationships within the data without explicit programming.

Hypothesis

There is a significant relationship between demographic characteristics, VR headset type, duration of experience, motion sickness rating, and immersion level in virtual reality environments, enabling the possibility of developing predictive models to estimate users' immersion level based on these variables.

Methods

- Data Loading
  - ➢ Obtain the dataset on Kaggle website and load it into a DataFrame in Jupyter Notebook.
- Data Exploration
  - ➢ Visualize the dataset to gain insights into its structure and characteristics.
  - ➢ Investigate the data types of each column (e.g., numerical, categorical).
  - ➢ Check for missing values to ensure data integrity.
- Data Preprocessing
  - ➢ Select relevant features that contribute significantly to the target variable and remove irrelevant features.
  - ➢ Encode categorical variables into numeric representations (e.g., one-hot encoding, dummy encoding)
- Model Training
  - ➢ Split the dataset into training and testing sets to train and evaluate the model's performance. The ratio of the train-test split is 80:20. This means that 80% of the data will be used for training machine learning models, and the remaining 20% will be used for evaluating the models' performance.
  - ➢ Train the model on the training data using the chosen algorithms (e.g., logistic regression, k-nearest neighbors, random forest, etc.)
- Model Evaluation
  - ➢ Assess the model's performance on the testing data.
  - ➢ Use appropriate evaluation metrics (e.g., accuracy for classification, mean squared error for regression) to measure how well the model generalizes to unseen data.
- Python Libraries: pandas, scikit-learn, seaborn, matplotlib, plotly.

Acknowledgements

- Project supported by Project RAISER, U.S. Department of Education HSI-STEM award number P031C210118.
- Faculty Mentor: Dr. Doina Bein of Computer Science Department at California State University, Fullerton.
- Peer Mentor: Gia Minh Hoang from Dr. Bein lab of Computer Science Department at California State University, Fullerton.

Results

Table 1. Original VR dataset

	UserID	Age	Gender	VRHeadset	Duration	MotionSickness	ImmersionLevel
0	1	40	Male	HTC Vive	13.598508	8	5
1	2	43	Female	HTC Vive	19.950815	2	2
2	3	27	Male	PlayStation VR	16.543387	4	2
3	4	33	Male	HTC Vive	42.574083	6	3
4	5	51	Male	PlayStation VR	22.452647	4	2

Table 2. Binary encoded gender and VR headset features

	Age	Duration	MotionSickness	Gender_Female	Gender_Male	Gender_Other	VRHeadset_HTC Vive	VRHeadset_Oculus Rift	VRHeadset_PlayStation VR
0	40	13.598508	8	0.0	1.0	0.0	1.0	0.0	0.0
1	43	19.950815	2	1.0	0.0	0.0	1.0	0.0	0.0
2	27	16.543387	4	0.0	1.0	0.0	0.0	0.0	1.0
3	33	42.574083	6	0.0	1.0	0.0	1.0	0.0	0.0
4	51	22.452647	4	0.0	1.0	0.0	0.0	0.0	1.0

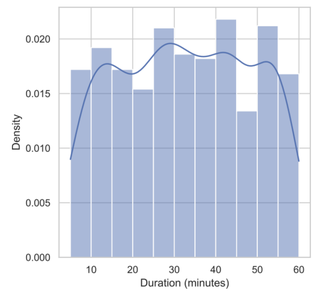

Figure 1. Distribution of Duration in Minutes

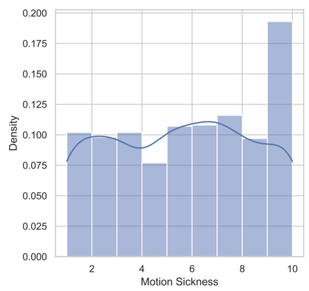

Figure 2. Distribution of Motion Sickness

On Figure 1 and Figure 2, the Kernel Density Estimate (KDE) curve provides an estimate of the probability density function, allowing us to observe the underlying shape and pattern of the distribution.

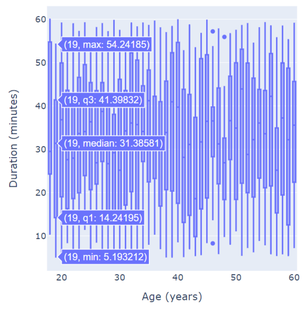

Figure 3. Age vs. Duration

A box plot displays the distribution of duration values for different age groups, allowing for easy comparison of central tendencies and spread across age categories.

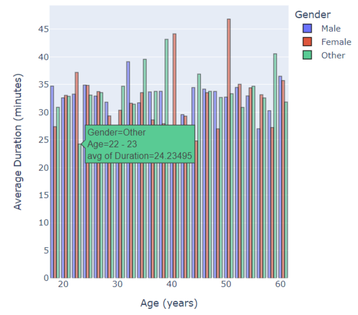

Figure 4. Average Duration by Age and Gender

The plot displays the average duration distribution for different age groups and genders using a grouped histogram.

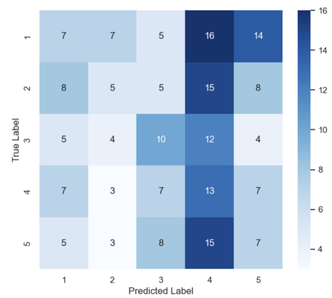

Figure 5. Confusion Matrix

The heatmap shows the confusion matrix with true immersion level (1-5) as row, predicted label as column, and cell values representing the count of instances for model performance evaluation.

Results Cont.

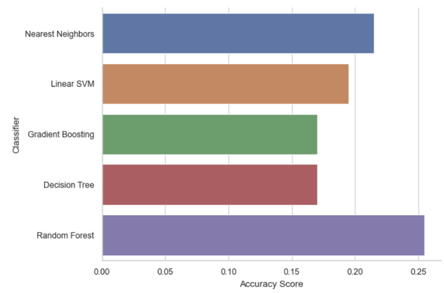

Table 3. Accuracy Comparison of Different Classification Models

	Classifier	Accuracy Score
0	Nearest Neighbors	0.215
1	Linear SVM	0.195
2	Gradient Boosting	0.170
3	Decision Tree	0.170
4	Random Forest	0.240

Figure 6. Accuracy Comparison of Different Classification Models

The bar plot (figure 6) visually compares accuracy scores of different models to identify the best performer. The DataFrame (table 2) lists model names and their accuracy scores, presenting an organized view of predictive performance, where higher scores indicate better results.

Conclusion

- The research has provided valuable insights into the factors influencing users' immersive experiences in virtual reality environments, even though the predictive models' accuracy scores were lower than expected.
- The findings suggest important implications for enhancing virtual reality experiences through personalized and optimized approaches based on age, gender, VR headset type, duration of experience, and motion sickness rating.

Future Work

The relationships between the features and the target variable demand a more substantial and diverse dataset to draw firmer conclusions. Further research and improvements in the modeling approaches may be necessary to achieve higher accuracy and ensure that users have engaging and satisfying interactions in the virtual environments.

References

- Aathi, K. (2023). *VR4 -> visualizations + EDA + model building*. Kaggle. https://www.kaggle.com/code/aathikm/vr4-visualizations-eda-model-building
- Nantasenamat, C. (2020). *Compare Machine Learning Classifiers in Python*. Retrieved August 2, 2023, from https://www.youtube.com/watch?v=QlNjjSge65Y&list=PLtqF5YXg7GLltQSLKSTnwCcHqTZASedbO&index=2.
- Nantasenamat, C. (2022). *Build your first machine learning model in Python*. Retrieved August 2, 2023, from https://www.youtube.com/watch?v=29ZQ3TDGgRQ.
- Joshi, A. (2023). *Virtual reality experiences*. Kaggle. https://www.kaggle.com/datasets/aakashjoshi123/virtual-reality-experiences
- Müller, A. C., & Guido, S. (2018). *Introduction to machine learning with python: A guide for data scientists*. O'Reilly Media.
- Raschka, S., & Mirjalili, V. (2019). *Python machine learning: Machine learning and deep learning with python, scikit-learn, and tensorflow 2*. Packt Publishing.